

Estudos Aplicados de Economia

Licenciatura de Economia - Disciplina de Opção

<http://www2.fe.uc.pt/~jasa> e perlopes@fe.uc.pt

FACULDADE DE ECONOMIA - UNIVERSIDADE DE COIMBRA, 2003-04

Representação de variáveis

Bibliografia

1. Bento Murteira, C. Silva Ribeiro, J. Andrade e Silva e Carlos Pimenta, *Introdução à Estatística*, McGraw-Hill, 2001, Lisboa
2. Bento Murteira, *Análise Exploratória de Dados*, McGraw-Hill, 1993, Lisboa (sobretudo Cap. 2,3 e 4)

Identificação do tipo de variáveis

Variáveis discretas e contínuas (stock e fluxo)

1. contínuas e a sua medida
2. discretas e a sua medida
3. discretas como contínuas, desde que possível
4. ... mas são diferentes

Não esquecer um princípio importante:

- métodos de análise não devem variar quando uma pequena fracção toma valores invulgares !

O uso de gráficos

Principal objectivo: criar sensibilidade ao aspecto global !

Caule e folhas:

- **caule**: dígito dominante
- **folha**: dígitos dominados
- exemplo: 3|06, corresponde a 3.0 e 3.6
- veja-se o caso das temperaturas médias e do crescimento entre 1990 e 2000 das *High Income economies*. Cálculos no Excel.
- profundidade: menor das ordens da contagem crescente e decrescente
- percepção de concentração e dos extremos
- número de linhas ?
 - Velleman sugeriu $2 \cdot \sqrt{20}$ arredondado à potência mais próxima de 10 : $2 \cdot \sqrt{20} = 8.9 \rightarrow 8$
 - cálculo das linhas: $19.4 - 11.4 = 8$, e como $\frac{8}{8} = 1$. Linhas de 1 grau.

Exemplo com temperaturas médias

B. Murteira, *Análise Exploratória de Dados*, p.16/7

				Profundidade				
Bragança	13,2	Guarda	11,4	1	20	1		
Viana	15,6	Miranda	13	2	19	2	N=20	
Braga	15,3	Bragança	13,2	3	18	3	Profund.	Unidad Décimas
V Real	14,1	V Real	14,1	4	17	4	1	11 4
Miranda	13	Viseu	14,4	5	16	5		12
Porto	15,5	Braga	15,3	6	15	6	3	13 0,2
Viseu	14,4	Porto	15,5	7	14	7	5	14 1,4
Guarda	11,4	Viana	15,6	8	13	8	9	15 3,5,6,8
Coimbra	16,6	Portalegre	15,8	9	12	9	(5)	16 0,2,6,7,7
C Branco	16	C Branco	16	10	11	10	6	17 1,4,5,6,7
Portalegre	15,8	Évora	16,2	11	10	10		18
Santarém	16,7	Coimbra	16,6	12	9	9	1	19 4
Lisboa	17,5	Santarém	16,7	13	8	8		Funchal
Évora	16,2	Setúbal	16,7	14	7	7	(5)	8,94427
Setúbal	16,7	Beja	17,1	15	6	6		
Beja	17,1	P Delgada	17,4	16	5	5		
P. Rocha	17,6	Lisboa	17,5	17	4	4		
Faro	17,7	P. Rocha	17,6	18	3	3		
P Delgada	17,4	Faro	17,7	19	2	2		
Funchal	19,4	Funchal	19,4	20	1	1		

O uso de gráficos ...

... Caule e folhas:

- ...
- cuidado: valores invulgares levam a concentrar informação !
- como reconhecer e que fazer com *outliers*?

Quadro de frequências

- variáveis discretas
 - numerosas observações
 - da frequência absoluta à relativa
- variáveis contínuas
 - definição das classes (intervalo de classe)
 - valores por classe e o mesmo que as discretas

Crescimento 1990-2000, W.D.R., WB

Crescimento

Suíça	0,7	1	23 [✓]	1		
Islândia	1	2	22	2		
Japão	1,3	3	21	3	1	0 7 Suíça
Luxemburgo	1,4	4	20	4	8	1 3,4,5,5,7,8
Alemanha	1,5	5	19	5	✓(10)	2 0,0,4,4,4,5,6,8,9,9
Itália	1,5	6	18	6	5	3 0,4,6
França	1,7	7	17	7	2	4 1
Suécia	1,8	8	16	8		5
Áustria	2	9	15	9		6
Bélgica	2	10	14	10	1	7 3 Irlanda
Dinamarca	2,4	11	13	11		
Grécia	2,4	12	12	12 ^{✓(10)}		
Espanha	2,4	13	11	11		
Reino Unido	2,5	14	10	10		
Portugal	2,6	15	9	9		
Finlândia	2,8	16	8	8		
Holanda	2,9	17	7	7		
Canadá	2,9	18	6	6		
N. Zelândia	3	19	5	5		
E.U.A.	3,4	20	4	4		
Noruega	3,6	21	3	3		
Austrália	4,1	22	2	2		
Irlanda	7,3	23	1	1		

O uso de gráficos ...

Representação gráfica

Diagrama de barras

- variáveis discretas

Histograma (diagrama de áreas)

- variáveis contínuas
 - base: intervalo de classe
 - altura: frequência relativa

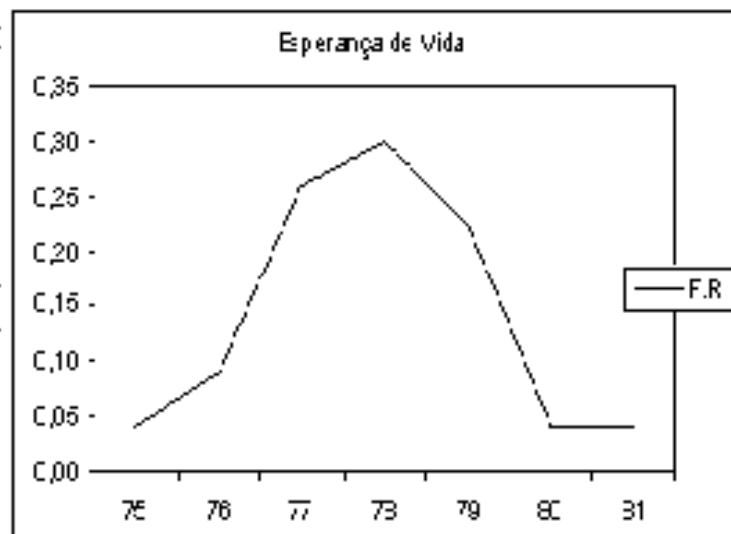
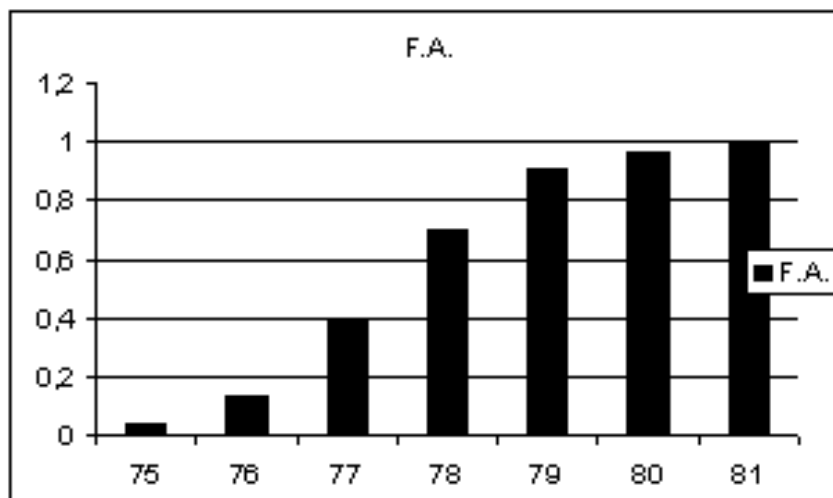
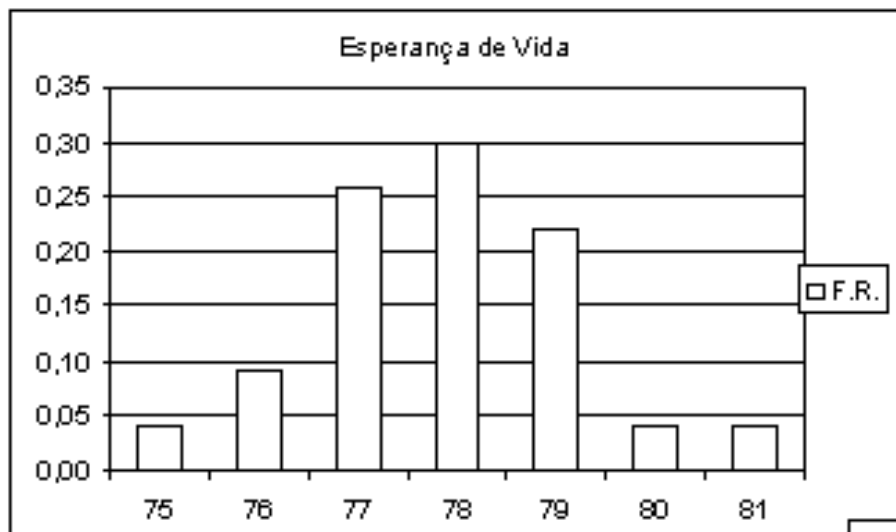
Polígono de frequências

- base: valor médio do intervalo de classe
- altura: frequência relativa

Esperança de vida, W.D.R., WB

Esp Vida		FrAb	FrRe		
Portugal	75	1	0,04		
Dinamarca	76				
Irlanda	76	2	0,09		
Finlândia	77				
Alemanha	77				
Luxemburgo	77				
N. Zelândia	77				
Reino Unido	77				
E.U.A.	77	6	0,26		
Áustria	78				
Bélgica	78				
Grécia	78				
Itália	78				
Holanda	78				
Noruega	78				
Espanha	78	7	0,3		
Austrália	79			75	0,043
Canadá	79			76	0,087
França	79			77	0,261
Islândia	79			78	0,304
Suécia	79	5	0,22	79	0,217
Suíça	80	1	0,04	80	0,043
Japão	81	1	0,04	81	0,043
		23	1		

Esperança de vida, ...



O uso de gráficos ...

Representação gráfica ...

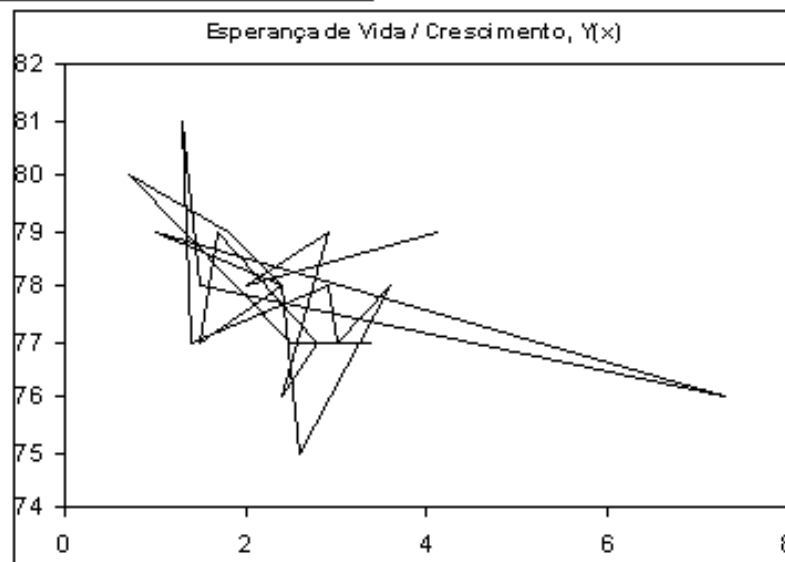
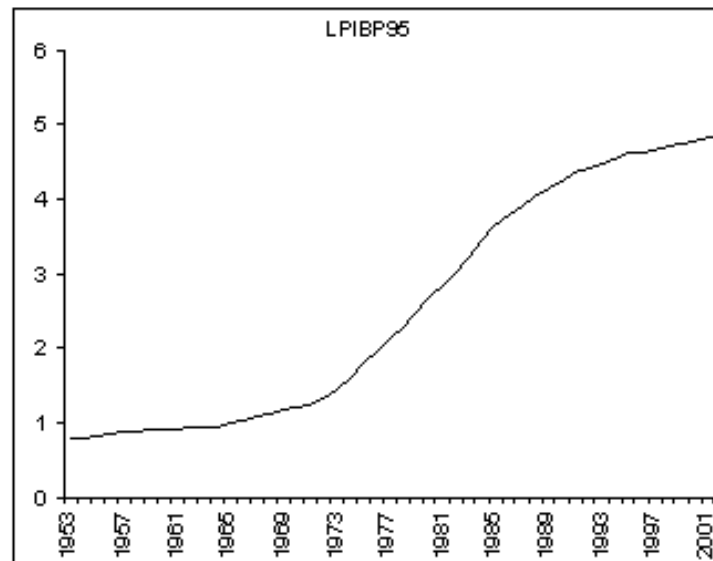
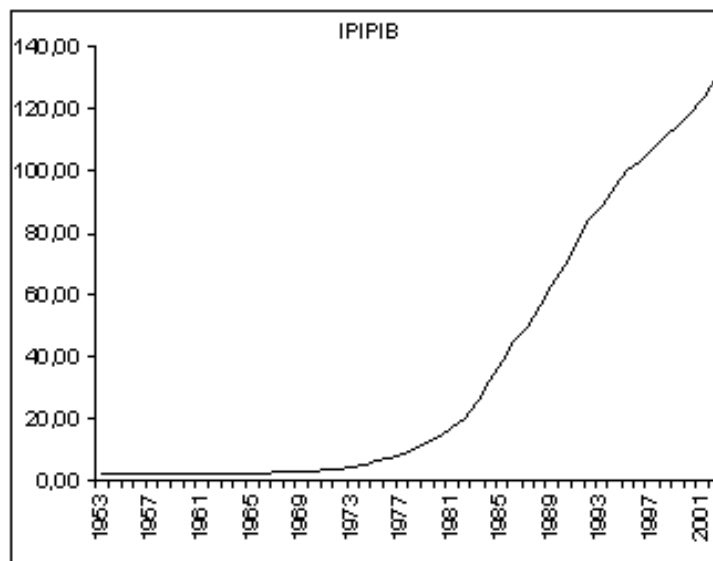
Cronograma de uma variável

- escala e dupla escala
- valores efectivos, logs ou índices

Diagrama de relação

- escalas
- unidades das variáveis

O uso de gráficos ...



Conceitos a reter

média, moda, mediana, desvio-padrão, coeficiente de dispersão e outliers

- outliers: erro humano ou fenómeno estranho. Regra: $3 \cdot \pm\sigma$
- desvio-padrão (N) e desvio-padrão corrigido ($N - 1$),

$$s = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}}$$

- assimetria. De Pearson $g = \frac{\bar{X} - Mod}{s}$

- valores distribuídos como uma Normal?

- Façamos $m_k = \frac{1}{N} \cdot \sum_{i=1}^N (X_i - \bar{X})^k$

- Skewness (enviesamento) $SK = \frac{m_3}{m_2^{3/2}}$

- Kurtosis $KU = \frac{m_4}{m_2^2}$

- Estatística de *Jarque-Bera* $J - B = N \cdot \left(\frac{(KU - 3)^2}{24} + \frac{SK^2}{6} \right) \sim \chi_2^2$

Conceitos a reter ...

correlação e estatística de Ljung-Box

- interprete: $SK = 0,14$, $KU - 3 = 0,41$ e $J - B = 0,233$ com $0,890$ de nível de significância.
- Coeficiente de correlação entre duas variáveis:

$$\rho_{x,y}(k) = \frac{\sum (x_t - \bar{x}) \cdot (y_{t-k} - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2 \cdot \sum (y_t - \bar{y})^2}}$$

- Ljung-Box (hipótese nula de não correlação) com $\sim \chi^2(\text{desf})$. L-B para desfasamentos de M_1 a M_2

$$Q = T \cdot (T + 2) \cdot \sum_{M_1 \leq j \leq M_2} \frac{\hat{\rho}_{x,y}^2(j)}{T - |j|}$$